图书情报工作
第 66 卷 第 20 期 2022 年 10 月

网络分析在数字人文中的应用框架*

■ 王林旭 王军
北京大学信息管理系 北京 100871 北京大学数字人文研究中心 北京 100871

摘 要:[目的/意义]网络分析作为有效的分析手段与可视化方式,是数字人文领域应用最广泛的方向之一。对网络分析在数字人文领域的应用进行系统总结与归纳,有助于数字人文研究人员快速明确网络分析的能力与局限,以便开展更深层次的研究实践。[方法/过程]采用内容分析法对发表在国际数字人文领域最有影响力的期刊和国际数字人文会议上近 5 年的文章进行梳理与总结,从研究问题、数据集、网络特性、网络分析指标 4 个方面进行梳理,最终提炼出网络分析方法在数字人文领域的应用框架。[结果/结论]该应用框架包括三类数据规模(单一文本、平行文本和语料库)、五大应用场景(角色分析网络、人物关联网络、话语空间网络、文本关联网络、文化主题网络)、两种指标规模(全局指标和局部指标)以及五类指标类型(构成、密度、中心度、派系和结构)。

关键词:网络分析 数字人文 社会网络分析 内容分析
分类号:G250
DOI:10.13266/j.issn.0252-3116.2022.20.006

1 引言

20 世纪末以来,由人文计算(humanities computing)发展而来的数字人文(digital humanities)带来了全球范围内知识生产范式的转型。2016 年后,数字人文在中国大陆也进入了加速发展阶段,网络基础设施建设和由研究问题驱动的数字人文研究之间的边界逐渐清晰,“方法共同体”逐步显现[1]。网络分析源自网络理论,不仅具有强大的分析能力,也是数据可视化的一种有效方式,近年来在数字人文领域得到广泛应用,成为取得实质性进展最多的方向之一。

网络是表征系统各部分间联系及交互模式的强大工具,社会网络(social network)是其中之一。社会网络分析(Social Network Analysis,SNA)关注参与者(即人),与数字人文研究的基本思路不谋而合。C. Wetherell 认为,虽然社会网络分析对数据量的要求十分苛刻,需要社会系统内所有成员之间不同社会交互行为的证据信息,但对史学分析而言,社会网络分析仍然有着巨大潜力[2]。社会网络分析主要依靠复杂网络(complex network)分析和可视化技术对社会网络节点之间的各类关系数据进行定量或定性分析,关注成对个体的属性,常见的属性有亲属关系、社会角色、情感、认知、行动、流量、距离和共现等。

数字人文领域进行社会网络分析可以为各种社会关系提供精确的量化分析,从而为构建理论模型和验证命题提供社会化例证,挖掘出更多隐含的社会关系和变化趋势。利用社会网络分析技术对特色资源进行分析和建模,关注的焦点是节点的获取与处理、关系与关系模式的确定;社会网络分析采用的方式和方法从概念上有别于传统的统计分析和数据处理方法,是探索数字化文史材料上人文问题的新工具。随着社会网络分析在数字人文研究中的逐步开展,人物不再是网络中的唯一节点,由文本、词语等元素构成的网络开始被学者们关注并在研究中使用。故本文所讨论的网络不限于仅由人物构成的社会网络。

为探究数字人文研究中网络分析的使用情况,明确方法的应用目标、能力范围与使用依据,本文使用内容分析法对相关文献进行梳理,提出一个 SNA 方法在数字人文领域中应用的结构框架,以便于从方法论层面熟悉网络分析方法,拓展网络分析的使用边界。基于此目的,本文要探索的核心问题是:①人文资料上使用网络分析可以开展怎样的研究;②在此类研究中,网络由哪些节点与联系构成;③在网络分析的过程中会选用哪些指标。

* 本文系国家自然科学基金国际重点合作项目“中国儒家学术史知识图谱构建研究”(项目编号:72010107003)研究成果之一。
作者简介:王林旭,博士研究生;王军,教授,通信作者,E-mail:junwang@pku.edu.cn。
收稿日期:2022-07-29 修回日期:2022-09-05 本文起止页码:50-61 本文责任编辑:杜杏叶

王林旭，王军. 网络分析在数字人文中的应用框架[J]. 图书情报工作,2022,66(20):50-61.

2 研究设计

2.1 文献获取

本文以数字人文领域最有学术影响力的期刊 *Digital Scholarship in the Humanities*(以下简称 *DSH*),以及数字人文国际联盟(Alliance of Digital Humanities Organizations)组织举办的数字人文大会(以下简称 DH 大会)作为主要数据来源,收集 2018 年至 2022 年 5 年收录的论文,筛选出使用网络分析对文本进行解构、分析与挖掘的研究。此外,考虑到 *DSH* 为语言学领域期刊,较多关注语言学、文学领域,本文将鼓励数字人文主题研究的历史领域期刊 *Historical Methods* 论文纳入其中。最终筛选得到 32 篇文章作为研究对象,如表 1 所示:

表 1 文章(研究对象)来源分布

来源	领域:2021 年影响因子(所在区)	论文数量/篇
Digital Scholarship in the Humanities(*DSH*)	LINGUISTICS:1.299(Q2)	24
DH Conference(DH)		6
Historical Methods: A Journal of Quantitative and Inter Disciplinary History	HISTORY:1.647(Q1)	2
《数字人文》	-	5
《数字人文研究》	-	3

近年来,国内数字人文的学术交流与出版活动日渐繁荣,清华大学与中华书局创办的《数字人文》和中国人民大学信息资源管理学院推出的《数字人文研究》都是数字人文领域的专业期刊(见表 1)。考虑到这两本期刊创刊至今均不足五年,本文未将二者纳入编码讨论的范围。但上述两本期刊中的相关研究对中文语境下的材料处理与方法应用有着重要参考价值,本文在对研究结果的讨论中也对其相关的 8 篇原创论文进行了总结与归纳。此外,欧洲的 *Journal of Historical Network Research* 是少有的聚焦网络分析在领域内使用的人文社科类期刊,自创刊以来,5 期均发表了历史、社会和政治科学对社会网络分析方法的高水平应用实践。前不久,其特刊"Beyond Guanxi"正式出版,旨在向欧美史学界介绍近年来中国历史网络的学术成果,为面向中文历史材料的网络分析提供了研究范例,故本文在讨论中也将其考虑在内。内容分析法建立在对文本内容的深度解读上,编码工作量大。上述相关文献汇集起来,已能达到本文的研究目的。

2.2 研究方法

文献内容分析法是定性与定量相结合的分析方法,它以定性的问题假设作为出发点,通过对文献内容的定量分析,找出能反映文献内容的本质又易于量化的特征[3]。特别对于容纳不同学科研究的数字人文领域来说,学者们往往具有不同的写作习惯,对于研究数据与方法的阐释方式也各不相同、详略不一,仅仅通过阅读得出方法论的使用现状并不容易。内容分析法可以将用语言表示的文献转换为用统计数字描绘的资料,并从统计数据中得出定性的最终结论,从而帮助本文对相关文献有更深刻、更精确的认识。就学术文献而言,内容分析对象的主要内容是题目、摘要、关键词、参考文献等书目信息,但这些并不足达到对研究问题的充分理解与研究方法的整体把握,故为确保编码范围的完整性,本文将文献全文作为分析对象。

首先基于对数字人文领域研究的理解,形成了包括研究问题、研究数据、网络数据、网络类型与网络分析指标的 5 个主要一级编码;其次,基于文章内容,对上述 40 篇文章进行了二级编码,如表 2 所示,编码结果汇总可见第 6 节内容。由于编码对象多为英文论文,故其中二级编码的名称选择在尽可能多地汇总相同原文描述后使用具有概括性的中文词语。

表 2 编码范畴

一级编码		二级编码
研究问题		角色人物功能,叙事模式/体裁,叙事内容,关联发现,人物派别/社区界限
研究数据		戏剧,史诗,信件,小说,记录报告等
网络数据	节点	人物,词语,文本
	关系	人物对话,人物共现,人物所属(派系/社区等),人物关注(主题等),词语共现,词语相邻,文本相似
网络类型		角色对话网络,人物关系网络,话语空间网络,文化主题网络,文本关联网络
网络分析指标		点、线、大小,线、邻域和密度,中心度、边缘性和中心势,成分、核心和派系,位置、集和和聚类,网络动力和网络变迁
其他方法		数据库构建,统计描述,文本聚类,文本相似计算,可视化展示

在对相关文献的编码基础上,本文按照数字人文领域网络分析研究开展的顺序进行分析,包括确认研究问题与来源文本(见第 3 节)、明确节点与边以构建网络(见第 4 节)、选用不同的指标对构建的网络进行分析(见第 5 节),并结合具体的文献或不在本次编码范围内却引起领域关注的高水平文章对编码结果进行补充说明。最后,本文将 *DSH*、DH 大会与 *Historical Methods* 中的相关文章与核心编码进行对应,以提供对重要文献的具体分类(见第 6 节),便于更直观地理解网络分析的应用体系。

3 常见来源文本与研究问题

3.1 来源文本

数字人文中的网络分析通常从定义特定的文本语料库开始,即从处理分析材料开始。语料库的构建是得出有意义的结论的关键[4],通过探讨来源文本的可分析维度与研究问题的关联并定义语料库,网络分析可以将定性的研究问题进行量化与可视化[5]。数字人文研究有着跨学科的天然特质。除了人文、艺术、社会学研究与数字人文研究高度相关外,计算语言学、应用计算机科学本身的研究也与数字人文研究有着紧密的联系[6]。这些研究上的重合使得很难从学科偏好上将数字人文研究分类。尽管在研究范式上各有不同,但在同类研究数据上的探索往往使得彼此的研究互为补充,单一文本分析、平行文本分析与基于语料库的分析 3 种研究思路是不同学科在同类数据集上进行尝试的普遍路径。

数字人文领域应用网络分析进行研究的主要材料有:包括戏剧、小说、诗歌、散文的文学作品,包括信件,历史人文档案的纪实档案,包括词典、百科全书等的解释性文档,以及宗教学、哲学典籍文本等。此外,社交媒体的使用数据也可以用于探索记忆的传承与文化的传播。如 II. M. A. Rhodes 利用网络分析对 YouTube 上民权运动家保罗·罗伯逊(Paul Robeson)相关视频的点赞、评论进行分析,以揭示保罗·罗伯逊是如何被记住[7]。

统一类型的文本常常有着较为相似的研究问题与路径。通过将研究问题进行归纳,并与研究数据进行对应,本文共得到了 5 类主要研究问题,分别是角色功能分析、人物关系分析与派别归纳、叙事模式与内容分析、人物关联发现、数据库可用性评价与展示。如表 3 所示:

表 3 应用社会网络分析的数字人文研究问题与对象

研究问题	研究对象
人物角色功能	戏剧/诗歌
人物关系与聚类分析	信件/历史档案/结构化主题数据库
叙事模式与内容	戏剧/小说/信件
关联发现	结构化主题数据库
数据库可用性评价与展示	–

3.2 研究问题

3.2.1 文学作品中角色功能分析

针对文学作品的网络分析是数字人文在此类研究中最主要的议题。斯坦福大学比较文学系弗兰科·莫雷蒂(Franco Moretti)教授提出的建立在对大量文本数据收集和统计分析基础上的“远读”(Distant Reading)在世界文学研究中影响深远,为小说社会网络分析提供了理论依据与初步构想。建立在对文学作品角色网络的关键节点(即核心角色)及其之间发生对话这一行为基础之上的“情节分析”(Plot Analysis)[8]关注文本中的空间与时间、网络社群集团的划分、核心人物的发现等与小说叙事研究紧密相关的几个方面,以及基于网络理论解构作品中人物对话与关系,从大容量、多人物文本中发现真正具有叙事潜能的关键角色。

随着更多具体的量化实践的展开,角色之间“关系”的界定成为了研究的难点与热点,也使得基于不同题材文学作品的研究对“关系”进行不同的定义。例如戏剧中的对话行为几乎代表戏剧行为的全部,L. Evalyn 等将莎士比亚戏剧中角色在对话中共现作为节点之间的联系,人物说话的次数作为边的权重,通过人物交流网络来探究喜剧、悲剧和历史剧的叙事模式[9];但在小说文本中,对话并非叙事的全部,一定量的转述也可以揭示人物关系的紧密程度,赵薇就在“《大波》三部曲”中通过区分直接引语和间接引语及其比重、区分对话情景来分析关键人物的叙事功能[10]。无论是小说、还是戏剧,从本质上说仍是一种语言艺术,将对话语言作为主要分析变量时,能够直观地明确作品中的话语权分布;此外,将其与传统文本细读更加精确、深入地结合起来,结合一些重要的统计概念在人物和情节研究中的表现来探讨相关角色或文本的叙事学功用,能够进一部挖掘作品的独特价值。此外,将网络分析应用于文学作品也有助于研究特定文本或作者对一段时间内其他作品的影响[11],并了解不同角色之间关系的演变[12-13]。

3.2.2 人物关系与聚类分析

角色功能分析是基于对文学作品中人物关系以及人物对话的把握。但在其他材料中,人物关系往往需要依靠覆盖一定时间范围材料的积累才可以获得,尤以同样关注人物的历史学研究更加显著。自 M. C. Alexander 等调查 15 世纪初美第奇(Medici)家族崛起[14]开始,考古学家和历史学家对网络分析方法的兴趣与日俱增[15]。对包括信件、历史人文档案等纪实档案的研究往往跨越较长的时间维度,如追溯到文艺复兴时期的通信网络[16]、揭示中世界苏格兰精英共同见证宪章起草的证人从属网络[17],都在百年的时间窗口下探寻人物关系。

得益于诸如中国历代人物传记资料库(China Bio-

graphical Database Project,CBDB)等半结构化的历史人物传记数据库的增多,基于大规模语料的人物关系分析成为了可能。严承希等建立了宋代人物的政治网络,阐述了宋代政治网络的关系演化模式[18];张力元等通过对比北宋和南宋的人物学术网络和政治网络,对宋朝士大夫思想转变进行了探析[19]。随着关系网络中人物越来越多,人物的聚集与离散也就愈加具有被解读的价值,个人在群体中的位置也成为了网络分析中的一大研究主题。例如 J. K. Ochab 等通过对 13 世纪中叶的捷克贵族共同见证宪章的人物网络进行聚类来分析某一具体时间节点前后贵族的聚集情况,即叛乱前后贵族的交往[20]。

3.2.3 文本叙事模式与内容分析

追溯社会网络分析的发展可以发现,社会网络分析深植于人类文化价值的符号体系中,20 世纪的结构主义语言学和符号学都给予社会网络研究生长的土壤。因此,从文本中提取可解释的符号,建立虚构形象的关系网络,发现文本背后潜在的叙事意图,成为另一种可行的分析范式。邱伟云等近年关注的报刊研究就通过构建词汇概念网络描绘了思想史的重要转变[21];此外,也有学者分别构建梁启超和陈独秀的著作中与国家建设主题相关的 30 个词语的共现网络,以考察二人著作差异与思想转变[22]。

当网络中的节点是单词(或字符)、词组或句子时,其相互关系构成了叙事空间,且该网络往往是有向的。对叙事空间的讨论包括不同体裁的文本叙事的差异、叙事单元的传递性以及相似性。但对叙述单元之间关系的常用度量较为单一,往往使用共现频率和互信息分数,即两个叙事单元一起出现在指定字符窗口内的可能性,忽略位置信息使得对其分析很难深入到内容中。Y. Yang 等引入阻力距离(resistance distances)来描述字符共现网络中的词汇关系[23]。此外,通过将叙事单元与邻近的其他单元进行连接,也可以对叙事内容进行进一步的探索。

3.2.4 人物关联发现与数据库可用性分析

人物关联发现研究主要得益于各种结构化的主题数据库,且通常没有具体研究问题指向,是非常典型的数据驱动研究,也常常作为数据库构建类研究的应用案例。

网络分析与可视化也往往是对数据集进行评价和研究潜力展示的重要手段。M. Levine 在介绍中国传记数据库(China Biographical Database, CBD)后通过两个历史网络分析案例说明 CBD 的有效性[24];P. K. Bol 通过网络分析对浙江梧州人士的血缘关系与学术联系变迁进行研究,进一步说明了 CBDB 在人物关系上的深入描述[25]。除依靠现成数据库,大部分数字人文学者仍然需要自建数据集用于研究。因此,许多数字人文研究的重点不再是分析方法与结论,而是数据集构建;网络分析自然也就成为了基于部分数据的尝试。J. Waxman 基于上千名巴比伦学者辩论内容的《塔木德》,构建了以学生、同事关系为主的学术关系和包含协作、分歧、反驳等 13 种学者间的互动的《塔木德》数据集,并使用网络分析与可视化对部分内容进行了分析[26];同样,M. R. Zambrano 等基于历史档案对厄瓜多尔现代建筑进行了讨论[27]。

4 网络构建

明确数据基础和研究问题后,定义网络中的节点与边、构建网络是此类研究的下一个步骤。通过相关文章中的网络类型与网络中节点和边的选择,如表 4 所示,可以发现,数字人文领域常见的"网络"有角色对话网络、角色共事网络、角色地点网络、人物关系网络、人物归属网络、话语空间网络、文本关联网络和文化主题网络。不难看出,网络类型与网络数据与研究问题也紧密关联,但本章不再局限于领域研究问题,而是单纯从网络中节点与边的内容对网络类型予以分析。

表 4 应用网络分析的数字人文研究中的网络类型与网络数据

网络类型	主要网络数据(节点 - 边 - 节点)
角色对话网络	人物角色 -(发生对话)- 人物角色
	人物角色 -(对话中提及)- 人物角色
角色共事网络	人物角色 -(参与同一事件)- 人物角色
角色地点网络	人物角色 -(到访同一地点)- 人物角色
人物关系网络	人物 -(场景共现)- 人物
人物归属网络	人物 -(所属)- 社区/派系
话语空间网络	语词 -(存在于同一文档)- 语词
	语词 -(场景共现)- 语词
	语词 -(左右相邻)- 语词
	语词 -(引用同一文本)- 语词
文本关联网络	文本 -(相似)- 文本
文化主题网络	人物/作品 -(关注)- 主题
	主题 -(与同一人物相关)- 主题

4.1 角色分析网络

与角色人物功能研究目标相对应的网络类型是角色对话网络、角色共事网络、角色地点网络,可以将上述三种网络归纳为对文本中的虚构角色进行分析的角

色分析网络。

角色的对话空间是此类网络中角色功能最显著的具化。角色对话网络中，节点往往是人物，关系是两个人物之间发生对话，并通过对话发生的次数为关系赋以权重。I. Pikkanen 将 1837 年至 1869 年在瑞典和芬兰出版的 4 部历史剧人物作为节点，人物之间发生过交流作为边，通过 4 部戏剧，即 4 个时期的交流网络的度量指标和可视化揭示了戏剧中的核心人物与叙事空间[28]。M. Kubis 在几百部小说的基础上分别构建了 19 世纪和 20 世纪角色对话网络，并将说话者与对话中提及到的人进行了区分，以发现两个世纪之间的小说的系统差异[29]。

角色共事网络是指人物角色共同参与同一事件。P. Jayannavar 等研究了参与同一事件的角色交互网络，以及互相认识的角色对彼此的观察网络[30]；许超等在《左传》标注语料的基础上将人物与事件实体通过共现网络表示出来，发现了春秋网络的小世界性[31]。

角色地点网络则是角色之间到访过同一地点。J. S. Y. Lee 等将出现在同一地点的人物关联了起来[32]；郭佳欣等不仅构建了《长安十二时辰》的“角色 – 地点”网络以解析人物集体的活动范围和场所，还通过“地点 – 地点”网络分析地点的转移和关联，其中边的权重与各节点之间联系的紧密程度相关。此外，还有将角色在同一场景下共现作为节点之间的联系[33]，本文也将其纳入角色地点网络之中，此时的地点可以视为虚构的“地点”。

4.2 人物关系网络

人物关系网络以基于事实与多重可参考材料的人物与人物之间的各种关系作为依据，主要有两类网络，第一类网络描述了包含人物与人物场景共现的直接关系网络；第二类网络描述了人物与其所属社区（或派系）的关系。

人与人的关系种类众多，亲属关系、学术关系、政治关系都是常见的分析维度。W. Shang 将《世说新语》中的人物作为网络节点，每个故事中人物的共现作为边，并通过将人物关系分为以表现两人之间有事实层面的直接积极关系和表现为认同、尊重的态度或交游、对答、学术交流、同僚、礼让、安慰等程度较轻的正面互动关系，并赋予不同的权重，以研究东晋贵族的社交网络[34]。即便不对人物关系进行明确，人物关联网络仍然具有适用性。C. Armand 将民国人物标记为节点，当 A 的传记中出现 B，则添加 A 到 B 的有向边，并检验人物是否基于共享属性（省级出身、教育等）形成具有结构的社会网络[35]。此外，在大规模数据的支持下，由人物之间互动而产生的关系，也能观察到独特的现象。M. J. Hill 等在书目数据库的基础上，构建作者的图书交易网络，以洞悉作者的历史观念及其背景[35]；李惠等将曾国藩和与他有书信往来的人关联在一起，从个人书信网络中挖掘名人群体的人脉网络，分析曾国藩不同时期的人物关系[37]。

人物与所属派系的关系网络常常通过现有所属关系探析新的人物集团的划分。J. J. Yoo 通过朝鲜王朝“诗人 – 诗歌社区网络”与“诗人 – 诗歌派系网络”发现跨派系的诗人交往和诗歌社区的聚集与重叠[38]。

此外，人物与其他内容的直接相联也对于确定“人物集团”有着重要参考价值，如 H. L. Xiong 将人物与职官进行关联，以试图描绘宋朝精英在“书院育才”的政策下的晋升之路[39]。

4.3 话语空间网络

话语空间网络是文字（或单词）和词语的网络，包括语词与语词的场景（或段落）共现网络、语词搭配网络、词语衍生网络。

分析语词在同一段落或句子的共现网络是最常见、也是最理想的探索给定文本集合整体语义结构的方式之一。共现意为“在恒定大小的上下文窗口内出现的成对术语之间的关系”[40]，窗口可以是任意长度的文本——一定数量的字符或单词、一个句子、一个段落或整个文档。窗口大小的设置往往取决于一个文档中预计有多少主题，小说习惯上把文本分割为较小的模块，以捕捉随着故事而展开的短暂主题；诗歌虽用词简洁，但每个词都有相当独立的语义表达，将每个句子看作一个窗口是一种颇为有效的方式；新闻等纪实报道类短文本总是围绕一个主题，基于整个文档进行语词共现分析更为明智。D. Gamermann 等将《小王子》中的单词作为节点，将单词共现在一个句子和一个句子的两个单词之间存在相同的第三个单词作为边，构建了两个单词共现网络，来分析不同语言的《小王子》译本中单词的使用[41]；C. Lee 基于 404 条新闻报导选出了 100 个高频词，并用高频词在同一文档中出现作为高频词之间的联系[42]。

语词搭配网络即语词与左右相邻语词构成的词组网络；语词衍生网络则是由具有共同构成（如英文中的词根）词语构成的网络，二者同是语料库语言学方法论的衍生。语料库语言学通过对大量的真实语言材料的分析和统计来建立自然语言处理规则，主要应用于教学、翻译、词汇、词义、词典和语法等领域，关注宏观（整

个语料库语言特征与文本类型)和微观(具体的词汇和语法等语言现象)两个方面[43],有助于对历史文本和话语的解释[44]。网络分析对宏观与微观上层面的研究予以了支持。H. Bonin 对英国议会的辩论内容进行分析,用"人民"(people)与"民主"(democracy 或 democra＊)二词左右相邻的词组成了词语搭配网络以对民主概念的转变进行分析[45];国内,邱伟云也将《新青年》中的与"政治"一词相关联的字词概念进行了关联[46]。词语衍生网络上,J. Longhi 汇编了 11 位 2017 年总统候选人官方账户发布的推文,并将出现在同一推文中的 islam 及其派生词进行关联,以强调话语分析在分析假新闻的时候至关重要[47]。

4.4 文本关联网络

文本关联网络关注的是段落和篇章,它们之间的联系是不同文本之间的相似性,相似性的大小决定了边的权重。为了构建这类网络,往往需要先对文本进行处理。G. Rotari 等把格林兄弟的每一封信件作为节点,将出现常用词共现的信件关联起来,并通过余弦增量应用于 50－150 个最常用词的标准分数(z-score)来确定信件之间的相似性(即边的权重),以分析格林兄弟写作风格的异同和他们写作风格在其职业或个人发展中的作用[48]。

文本共识网络的节点同样由文本或文本样本表示,节点之间的连结可通过两种方式建立。第一种方式通过计算样本之间的文本距离,为每个节点建立一个与其最近节点的最强链接和两个与其次近节点的较弱链接;第二种方式是对不同范围的常用词重复上述过程多次,并且将每次迭代产生的网络以类似于共识树分析的方式组成单个共识网络。C. Lee 基于英语小说和韩语小说构建了作者和翻译者共识网络,以识别不同作者和翻译者的文本风格[49]。

4.5 文化主题网络

文化主题网络不同于上述几类网络,它强调社会互动与关系能够以某种方式调节各类社会现象的结构与能动性。1990 年纽约学派将网络与文化之间的联系区分为四种研究:作为文化渠道的网络、作为形塑文化的网络、文化形式的网络和文化互动网络[50]。但在具有强实践性的数字人文研究中,社会学领域的详细划分还尚未涉及。在这里,文化常常被解构为概念、趋势,或外化为物品上的叙事属性;通过聚类以挖掘这些元素的聚集与主题。L. Giagnolini 等在研究关于艺术史学家丰富的传记档案基础上,构建了"历史学家－关注主题"网络、共同工作于同一机构的"历史学家－历史学家"网络,来发现艺术史社区中的关系,进一步发现文化的流动[51];S. Milonia 等通过"歌曲－歌曲"网络观察音乐模仿如何揭示中世纪浪漫抒情诗的相互关联的本质[52]。

5 网络分析指标与工具

按照不同的划分依据,网络分析指标有不同的分类。依据整体网络和局部网络的划分,可以将相关的指标分为全局指标和局部指标。但在具体使用的时候,这两者往往会同时使用,因此依据描述的目的可将这些指标划分为五类描述指标。

5.1 全局属性和局部属性

网络集中度测量的历史可以追溯到半个多世纪以前,但与更为客观和可靠的视觉解释不同,网络的度量结果十分需要依据上下文解读[53]。

如表 5 所示的全局属性的统计分析可以一次比较多个维度的网络[54]。例如,比较相同类型的不同网络的节点和边的数量[55]是一种可直接解读排名的工具。除此之外,网络密度(与节点相关的边数)与测量平均路径长度在分析人物网络时十分有效。

表 5 社会网络分析中的全局属性和局部属性

属性类型	包含指标
全局属性	节点数量(graph size-nodes)
	边数量(graph size-edges)
	网络密度(density)
	网络直径(diameter)
	平均路径长度(average path length)
	连通性(connectedness)
	集群/社区(clusters/communities)
	全局/平均集聚系数(clustering coefficient)
局部属性	某一节点的连通性(connectivity-degree)
	中介中心性(betweenness)
	接近中心性(closeness)
	特征向量中心(eigenvector centrality)
	局部集聚系数(local clustering coefficient)
	最短路径(shortest path)
	派(cliques)

局部度量中,度(相邻节点的数量)是最简单的中心性指标,且在 20 世纪 50 年代末至 20 世纪 70 年代初,该指标是唯一系统使用的指标,随后才发展出更加多样化的度量[56]。它的简单性使得其的解读十分明确,例如在文学网络中,度计算的可以是一个人物与另一个人物说话的次数。中介中心性(betweenness centrality)的概念打破了网络"中心"可能由什么组成的概

念，在揭示连接不同子网络的能力上很受欢迎。中介性与流通概念紧密相关，它通过计算最短路径以检测能够打开或锁定网络某些部分的中间“桥梁”或“关键通道”。因此，根据应用场合，这些位置既是网络中最具权力的位置，换言之，也是脆弱的位置。接近中心性（closeness centrality）允许节点在具有一定密度、且尚未划分派的网络中相当平均地分布，并且可以在“中心”和“外围”的概念中转化。特征向量中心性（eigenvector centrality）的解释相对困难许多，并且在很大程度上取决于节点周围的结构，但其在尝试分析图中节点的层次结构时特别有用[57]；同时特征向量中心性也是 PageRank 算法的依据。

5.2 5 类网络分析指标

在描述网络特征时，即可以描述每个点的特征，又可以把该网络看成一个整体，描述整个网络的特征。本文通过对相关文章进行编码，将上述指标划分为 5 个层次——构成、密度、中心度、派系和结构，相关论文对上述指标的使用频率如表 6 所示：

表 6 网络分析指标应用频率

网络分析指标	使用频率/%
构成（点、边、大小）	68.8
密度（线、邻域和密度）	46.9
中心度（中心度、边缘性和中心势）	50.0
派系（成分、核心和派系）	46.9
结构（位置、集和和聚类）	21.9

注：一篇文章可能会使用多种指标，故各类分析指标使用频率之和大于 1

（1）构成。点、边、大小是对网络中包含了多少个节点、以及一个节点关联了多少其他的节点的统计，是基于网络构成予以描述的最常见指标。研究者可以通过对个别节点的解释与网络可视化对特别的节点进行描绘。

（2）密度。本次编码的论文中，几乎 47% 的研究都对线、邻域和密度进行了分析，即对网络空间大小进行了进一步描述，故本文使用最能表示此类节点的密度一词予以概括上述三类指标。网络密度、网络直径、平均路径长度和某一节点的连通性是分析的核心依据。

（3）中心度。半数的研究都探讨了中心度、边缘性和中心势。在对每个点的特征进行描述时，计算与该点有直接联系的其他点有多少是一种重要的分析角度，这个值就是“中心度”；在对整体网络的特征进行描述时，通过“中心势”来反映该网络在多大程度上围绕着一个点而建立，即一个网络图向核心靠拢的程度。“中心度”和“中心势”都是对节点能力的量化视角，中介中心性、接近中心性、特征向量中心是衡量上述视角的关键。

（4）派系。成分、核心和派系是关于网络中子群的研究。尝试发现网络可以分为多少类派、聚类（clusters）、成分（componcents）、核（cores）、圈子（circlcs）等都是在挖掘网络中的子图，而子图的依据是点的关联性与关系的紧密程度。这些子图往往为研究者们提供节点分类的参考。

（5）结构。位置、集和和聚类的分析包括点的结构对等性、规则对等性、聚类、块模型等内容。对节点位置的探讨往往已经超越了图论的原则，是对集合的结构分析；当被应用于数字人文研究时，往往能够使得对个案的探究衍生到同类情况的判定。

5.3 网络分析工具

被学界和行业广泛认可并使用的网络分析工具有许多，Neo4j、Cytoscape、Gephi、Pajek、JUNG 等工具都在知识发现、信息融合、可扩展性和可视化方面有着不错的表现[58]。Gephi，一个在 NetBeans 平台上用 Java 编写的开源网络分析和可视化软件包，由于其在可视化方面出色的表现能力，被数字人文学者广泛使用。但由于 Gephi 在对网络指标计算支持上的薄弱，通常需要搭配 UCINET 这样的社会网络分析软件或 R、Python 这样的程序语言。

随着越来越多拥有人文学科背景的研究人员介入数字人文领域，许多数字人文研究平台也为不具有计算机编程习惯的学者提供了网络可视化与分析工具，如斯坦福大学的在线网络可视化工具 Palladio[59]，能够让用户轻松上传数据以展开网络分析；DocuSky 数字人文研究平台更是提供了具有 4 种分析方式的中文文本风格网络分析[60]。

6 文献分类

在了解了常见研究问题、文本数据、网络类型与分析指标后，本文得出了数字人文研究使用网络分析进行实践的应用框架；通过相关文章与核心编码对应，即对重要文献的具体分类，一方面可以便于更直观地理解文献详情，另一方面也验证了上述讨论的内容的适用性与覆盖面。来源文本的特征虽是进行量化分析的根本依据，但其种类的繁多与差异并不影响对方法论的探讨；而数据规模是决定量化方法使用的上限。故本文将相关文献区分为单一文本分析、平行文本分析与基于

语料库的分析三类研究。具体文献分类见表7。

6.1 单一文本分析

在对单一文本分析时,对所构建网络的阐释至关重要,往往需要结合现有研究与材料产生背景综合判定。由于数据量较少,网络中的节点与边也相对较少,通常从点、边、大小来展开的网络可视化分析和从线、邻域和密度,中心度、边缘性与中心势两个层面展开描述性分析已足够。此外,结合多种分析方法对单一文本进行深入挖掘也是此类研究展开的必要条件。V. H. Masías 等在加权网络的基础上使用 K-means 聚类对《罗密欧与朱丽叶》中的角色进行分析,其中仅选用了接近中心性作为加权网络的重要指标[61]。

6.2 平行文本分析

两个相似的文本或同一作品的不同语言翻译版本对比分析都可以看作是基于平行文本的分析。随着网络中节点和边的数量增多,派系(或凝聚子群)的生成与描述成为了对比分析的重要参考。Y. Fang 等通过《爱丽丝梦游仙境》的三个中文译本中的话语空间网络凝聚子群分析来探讨译者的风格[62]。

6.3 语料库分析(包括双语料库对比分析)

无论是作者自己构建语料库还是基于现有数据库进行研究,充足的数据使得网络分析方法能够有更多的施展空间。对比分析依然是常用的手段,T. Y. Lim 等将由普拉斯的224首诗歌作品所构建的话语空间网络与由塞克斯顿的280首诗歌作品构建的话语空间网络进行对比,通过重要节点、节点连通量、中介中心性、特征向量中心性、凝聚子群等各层面指标观察人称代词与情感词网络,此外还使用了结构主题模型来识别代表每个主题的一组词[63]。

表7 编码文献分类

分析特征		文献来源	分析层次					网络类型				
			构成	密度	中心度	派系	结构	角色关联网络	人物关系网络	话语空间网络	文本关联网络	文化主题网络
单一文本分析		H. A. Algee-Hewitt[64]	*							*		
		V. H. Masías 等[61]			*			*				
		M. Zhitomirsky-Geffet 等[65]	*	*					*			
		J. Waxman[26]	*						*			
平行文本分析		C. Ruegg 等[66]	*	*	*	*		*				
		D. Gamermann 等[41]		*		*				*		
		Y. Fang 等[62]	*			*	*			*		
语料库分析	对比分析	L. Evalyn[9]		*	*			*				
		M. Kubis 等[67]	*	*		*		*				
		G. Rotari 等[48]	*				*				*	
		C. Lee[68]	*			*					*	
		T. Y. Lim 等[63]	*	*	*	*				*		
	整体分析	M. J. Hill 等[36]			*	*			*			
		W. Li[69]		*	*				*			*
		T. Haider 等[70]			*	*	*		*			
		J. J. Yoo[38]			*	*			*			
		M. T. Santa María 等[33]	*		*	*		*				
		I. Pikkanen[28]	*		*			*				
		M. Kubis[29]		*	*	*	*	*				
		C. Jackson[17]		*	*				*			
		J. K. Ochab[20]	*	*	*	*			*			
		A. La Parra-Pereza 等[71]	*			*			*			
		D. M. Brown 等[72]	*				*		*			*
		S. Milonia 等[52]		*	*				*			*
		C. Lee[42]	*							*		*
		J. Ledolter 等[73], H. Bonin[45]	*	*						*		
		M. R. Zambrano 等[27]	*				*			*		
		J. Longhi[47]	*							*		
		Y. Yang 等[23]	*	*			*			*		
		L. Giagnolini 等[51]	*									*
		H. M. A. Rhodes[7]			*	*						*

7　未来挑战

近年来,随着"新文科"建设推进,数字人文在中国发展迅速,带来了材料和议题的拓展,也为人文社科领域注入了新活力。本文聚焦于特定方法的应用,发现国内外的数字人文研究虽不设边界,但也充满挑战,往往历时漫长,难以一蹴而就。即便研究者在进行实践探索时有着许多需要注意的内容,如理论和标准的探讨、研究的"效度"与"信度",但仍然有一些具体的问题值得我们考虑,它们作为当下研究的难点与挑战为我们未来研究提供了具体指引。

7.1　数据复用

近年来,使用网络分析的数字人文研究越来越多,网络分析范式在方法论层面的实践却没有逐步深入,只是通过将其应用于不同的材料与文本中解决类似的问题。从论文的篇章结构也能看出,即便是方法实践类研究,研究者们仍然将精力较多地放在了数据集构建上。因此,如何利用并共享现有的数据集以促进基于更多材料、更深层次研究问题的探索以及网络分析范式的拓展成为了数字人文领域未来的挑战,同时这也将成为数字人文研究边界的又一次拓展。让人欣慰的是,越来越多的半结构化或结构化主题数据集的开放使得数字人文学者凭一人之力也可以完成横跨不同时期的研究课题。中央华盛顿大学的 M. Levine 就在中国传记数据库(CBD)数据支持上利用主成分分析、地理信息可视化和一模网络结构分析来解释中国革命领袖的经历[74]。

7.2　多模网络

所有节点都为同一类型的网络称为"一模网络"(1-mode network);节点分为两类的网络称为"二分网络"(bipartite network)或"二模网络"(2-mode network);包含更多类别节点的网络则称为多模网络(multi-mode network)。由于网络分析工具的限制,许多学者不得不将二模网络拆分为多个一模网络以进行分析,通过对比得出结论;面向多模网络的分析更是少之又少。以较为普遍的成分、核心和派系来说,多模网络的派系发现主要有基于主题模型的方法、基于排序和聚类相结合的方法、基于数据重构的方法和基于降维的方法[75]。这些方法无一都对数字人文研究提出了更多的要求,但也为其带来了更多可能性。

8　结语

随着数字人文研究的兴起,数据与方法作为两类领域基础设施被频繁提起,而社会网络分析作为有效的量化与可视化手段取得了较多的成果。但国内外尚未见到对该方法在数字人文领域应用的系统化总结与归纳。故文本通过内容分析法,在领域期刊与会议的 40 多篇论文的基础上对此类方法的应用情况予以分析,为数字人文领域应用网络分析方法提供一个指导框架。

通过对国内外数字人文领域的网络分析研究进行梳理,本文得到了 SNA 在数字人文领域应用的结构框架,该框架从使用数据、研究问题、网络类型和分析、数据规模几个方面对网络分析的领域应用进行概括。其中,较为常见的研究问题可分为 5 类,分别是角色功能分析、人物关系分析与派别归纳、叙事模式与内容分析、人物关联发现、数据库可用性评价与展示;同时,根据节点与边的不同可进一步将"网络类型"总结为角色分析网络、人物关联网络、话语空间网络、文本关联网络和文化主题网络;此外,通过两种指标规模(全局指标和局部指标)和 5 个层面的分析指标(构成、密度、中心度、派系和结构),本文得以将研究实践的分析过程与研究问题关联,进一步丰富该框架的解释性。该框架不仅有助于理解 SNA 对数字人文研究的可用性、贡献等,还为研究人员找到适合的挖掘和分析角度,在方法论层面的探索提供支持。

参考文献:

[1] 赵薇. 数字时代人文学研究的变革与超越——数字人文在中国[J]. 探索与争鸣,2021(6):191-206, 232-233.
[2] WETHERELL C. Historical social network analysis[J]. International review of social history, 1998, 43(S6): 125-144.
[3] 邵作运,李秀霞. 引文分析法与内容分析法结合的文献知识发现方法综述[J]. 情报理论与实践,2020,43(3):153-159.
[4] RANDI R. Building a corpus: what are the key considerations?[M]//The Routledge handbook of corpus linguistics. London: Routledge, 2010.
[5] PAINTER D T, DANIELS B C, JOST J. Network analysis for the digital humanities: principles, problems, extensions[J]. Isis, 2019, 110(3): 538-554.
[6] LUHMANN J, BURGHARDT M. Digital humanities-a discipline in its own right? an analysis of the role and position of digital humanities in the academic landscape[J]. Journal of the Association for Information Science and Technology, 2022, 73(2): 148-171.
[7] RHODES II M A. Paul Robeson's place in YouTube: a social spatial network analysis of digital heritage[J]. Digital scholarship in the humanities, 2019, 34(1): 174-188.
[8] MORETTI F. Network theory, plot analysis[M]. London: Verso Books, 2013.
[9] EVALYN L, GAUCH S, SHUKLA M. Analyzing social networks

of xml plays: exploring shakespeare's genres[C/OL]. [2022－08－22]. https://dh2018. adho. org/analyzing-social-networks-of-xml-plays-exploring-shakespeares-genres/.

[10] 赵薇. 社会网络分析与"《大波》三部曲"的人物功能[J]. 山东社会科学,2018(9):50－64.

[11] SO R J, LONG H. Network analysis and the sociology of modernism[J]. Boundary 2, 2013, 40(2): 147－182.

[12] AGARWAL A, CORVALAN A, JENSEN J, et al. Social network analysis of alice in wonderland[C]// Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature. Montréal: Association for Computational Linguistics, 2012: 88－96.

[13] MAC CARRON P, KENNA R. Universal properties of mythological networks[J]. Europhysics letters, 2012, 99(2): 28002.

[14] ALEXANDER M C, DANOWSKI J A. Analysis of an ancient network: personal communication and the study of social structure in a past society[J]. Social networks, 1990, 12(4): 313－335.

[15] LEMERCIER C, ZALC C. Quantitative methods in the humanities: an introduction [M]//GOLDHAMMER A. Charlottesville: University of Virginia Press, 2019.

[16] AHNERT R, AHNERT S E. Metadata, surveillance and the Tudor State[J]. History workshop journal, 2019, 87: 27－51.

[17] JACKSON C. Using social network analysis to reveal unseen relationships in medieval Scotland[J]. Digital scholarship in the humanities, 2017, 32(2): 336－343.

[18] 严承希,王军. 数字人文视角:基于符号分析法的宋代政治网络可视化研究[J]. 中国图书馆学报,2018,44(5):87－103.

[19] 张力元,王军. 基于社会网络动力学的两宋学术和政治体系比较分析[J]. 情报工程,2020,6(1):34－49.

[20] OCHAB J K, ŠKVRŇÁK J, ŠKVRŇÁK M. Detecting Ottokar II's 1248－1249 uprising and its instigators in co-witnessing networks [J/OL]. Historical methods: a journal of quantitative and interdisciplinary history, 2022: 1－20 [2022－08－22]. https://doi.org/10.1080/01615440.2022.2065397.

[21] 邱伟云. 词汇、概念、话语:数字人文视野下中国近代"美"之观念的建构与再现[J]. 艺术理论与艺术史学刊,2019(1):182－223.

[22] CHAO A S, LIU Z, LI Q. Network of words: a co-occurrence analysis of nation－building terms in the writings of liang qichao and chen duxiu[J/OL]. Journal of historical network research, 2022, 5(1) [2022－08－22]. https://doi.org/10.25517/jhnr.v5i1.122.

[23] YANG Y, SONG Y. Exploring the similarity between Han's and non-Han's Yuan poetry: resistance distance metrics over character co-occurrence networks [J]. Digital scholarship in the humanities, 2022, 37(3): 880－893.

[24] LEVINE M. Biography for historical analysis: a chinese biographical database[J/OL]. Journal of historical network research, 2021, 5(1) [2022－08－22]. http://jhnr.uni.lu/index.php/jhnr/article/view/124.

[25] BOL P K. From kinship to collegiality: changing literati networks, 1100－1400 [J/OL]. Journal of historical network research, 2021, 5(1) [2022－08－22]. http://jhnr.uni.lu/index.php/jhnr/article/view/121.

[26] WAXMAN J. A graph database of scholastic relationships in the Babylonian Talmud [J]. Digital scholarship in the humanities, 2021, 36(S2): ii277－ii289.

[27] ROSA ZAMBRANO M, GONZALEZ M, MEDINA A, et al. Modern architecture in the professional discourse: analysis of the Architectural Biennial of Quito's 1976－92 archive using bipartite networks [J]. Digital scholarship in the humanities, 2022, 37(3): 894－909.

[28] PIKKANEN I. The metrics and poetics of historical drama: the dramatis personae of a premodern revolt in early nineteenth-century Finland[J]. Orbis litterarum, 2019, 74(5): 311－339.

[29] KUBIS M. Quantitative analysis of character networks in Polish 19th-and 20th-century novels[J]. Digital scholarship in the humanities, 2021, 36(S2): ii175－ii181.

[30] JAYANNAVAR P, AGARWAL A, JU M, et al. Validating literary theories using automatic social network extraction[C]//Proceedings of the fourth workshop on computational linguistics for literature. Denver: Association for Computational Linguistics, 2015.

[31] 许超.《左传》的语言网络和社会网络研究[D]. 南京:南京师范大学,2014.

[32] LEE J S Y, YEUNG C Y. Extracting networks of people and places from literary texts[C]//Proceedings of the 26th Pacific Asia conference on language, information, and computation. Bali: Faculty of Computer Science, Universitas Indonesia, 2012.

[33] SANTA MARÍA M T, CALVO TELLO J, JIMÉNEZ C M. Existe correlación entre importancia y centralidad? evaluación de personajes con redes sociales en obras teatrales de la Edad de Plata[J]. Digital scholarship in the humanities, 2021, 36(S1): i81－i88.

[34] SHANG W, SANG Z. Solidity in a turbulent flow: the social network of aristocratic families in the Eastern Jin Dynasty (317－420 CE)[J/OL]. Journal of historical network research, 2022, 5(1) [2022－08－22]. http://jhnr.uni.lu/index.php/jhnr/article/view/126.

[35] ARMAND C, HENRIOT C. From textual to historical networks: Social relations in the bio－graphical dictionary of republican China [J/OL]. Journal of historical network research, 2021, 5(1) [2022－08－22]. http://jhnr.uni.lu/index.php/jhnr/article/view/117.

[36] HILL M J, SÄILY T. Patterns of early modern authorship: using metadata as historical record[C/OL] Utrecht: DH, 2019[2022－08－22]. https://researchportal.helsinki.fi/sv/activities/patterns-of-early-modern-authorship-using-metadata－as-historical-.

[37] 李惠,侯君明,陈涛,等. 含珠与萤光:古代书信网络的衍化研究[J]. 数字人文,2020(1):109－117.

[38] YOO J J. Poetry in action: networks of literary communication and the cultural leverage in the eighteenth-century Seoul[J]. Digital scholarship in the humanities, 2021, 36(3): 766－781.

[39] XIONG H L. Path toward the top leadership: a network analysis of

the civil service system in the early southern song (1131 – 1164) [J/OL]. Journal of historical network research, 2022, 5(1) [2022 – 08 – 22]. http://jhnr. uni. lu/index. php/jhnr/article/view/98.
[40] G RAYSON S, WADE K, MEANEY G, et al. The sense and sensibility of different sliding windows in constructing co-occurrence networks from literature[C]//International workshop on computational history and data-driven humanities. Cham: Springer, 2016: 65 – 77.
[41] GAMERMANN D, MORET-TATAY C, NAVARRO-PARDO E, et al. The small-world of 'Le Petit Prince': revisiting the word frequency distribution [J]. Digital scholarship in the humanities, 2017, 32(2): 301 – 311.
[42] LEE C. How are 'immigrant workers' represented in Korean news reporting? – a text mining approach to critical discourse analysis [J]. Digital scholarship in the humanities, 2019, 34(1): 82 – 99.
[43] 徐彤阳,王霞. 语料库语言学视域下数据驱动的数字人文研究——以《数字人文季刊》为例[J]. 图书馆论坛,2021,41(10):90 – 99.
[44] MCENERY A, BAKER H. Corpus linguistics and 17th-century prostitution: computational linguistics and history [EB/OL]. [2022 – 09 – 01]. https://www. bloomsburycollections. com/book/corpus-linguistics-and-17th-century-prostitution-computational-linguistics-and-history.
[45] BONIN H. From antagonist to protagonist: 'democracy' and 'people' in British parliamentary debates, 1775 – 1885 [J]. Digital scholarship in the humanities, 2020, 35(4): 759 – 775.
[46] 邱伟云. 论数字人文研究中可视化数据的意义与价值——以数字概念史研究为例[J]. 文艺理论与批评,2020,(2):23 – 29.
[47] LONGHI J. Mapping information and identifying disinformation based on digital humanities methods: from accuracy to plasticity [J]. Digital scholarship in the humanities, 2021, 36(4): 980 – 998.
[48] ROTARI G, JANDER M, RYBICKI J. The Grimm Brothers: a stylometric network analysis[J]. Digital scholarship in the humanities, 2021, 36(1): 172 – 186.
[49] LEE C. Do language combinations affect translators' stylistic visibility in translated texts? [J]. Digital scholarship in the humanities, 2018, 33(3): 592 – 603.
[50] 刘军. 社会网络分析手册 上[M]. 重庆:重庆大学出版社, 2018.
[51] GIAGNOLINI L, DAQUINO M, MAMBELLI F, et al. Exploratory methods for relation discovery in archival data [J/OL]. Digital scholarship in the humanities, 2022[2022 – 08 – 18]. https://doi. org/10. 1093/llc/fqac036.
[52] MILONIA S, MAZZAMURRO M. Temporal networks of 'Contrafacta' in the first three troubadour generations[J/OL]. Digital scholarship in the humanities, 2022[2022 – 08 – 18]. https://doi. org/10. 1093/llc/fqac018.
[53] GRANDJEAN M, JACOMY M. Translating networks: assessing correspondence between network visualisation and analytics [C/OL]. Utrecht: DH, 2019[2022 – 08 – 18]. https://halshs. archives-ouvertes. fr/halshs-02179024/.
[54] TANG M C, CHEN K. A cross-language comparison of co-word networks in digital library and museum of buddhist studies [C/OL]. Montreal: DH. 2017[2022 – 08 – 18]. http://buddhism. lib. ntu. edu. tw/FULLTEXT/JR-OTHERS/others575593. pdf.
[55] TRILCKE P, FISCHER F, GOBEL M, et al. Theatre plays as ´small worlds? network data on the history and typology of german drama, 1730 – 1930[C/OL]. Jagiellonian University & Pedagogical University, Kraków: DH. 2016[2022 – 08 – 18]. https://publications. hse. ru/en/chapters/196225767.
[56] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social networks, 1978, 1(3): 215 – 239.
[57] PIPER A, ALGEE-HEWITT M, SINHA K, et al. Studying literary characters and character networks[C/OL]. [2022 – 08 – 18]. https://dh2017. adho. org/abstracts/103/103. pdf.
[58] CAMACHO D, PANIZO-LLEDOT Á, BELLO-ORGAZ G, et al. The four dimensions of social network analysis: an overview of research methods, applications, and software tools[J]. Information fusion, 2020, 63: 88 – 120.
[59] PALLADIO. Visualize complex historical data with ease[EB/OL]. [2022 – 08 – 18]. http://hdlab. stanford. edu/palladio/.
[60] TU H C, HSIANG J, HUNG I M, et al. Docusky, a personal digital humanities platform for scholars[J]. Journal of Chinese history, 2020, 4(2): 564 – 580.
[61] MASÍAS V H, BALDWIN P, LAENGLE S, et al. Exploring the prominence of Romeo and Juliet's characters using weighted centrality measures[J]. Digital scholarship in the humanities, 2017, 32(4): 837 – 858.
[62] FANG Y, LIU H. Seeing various adventures through a mirror: detecting translator's stylistic visibility in Chinese translations of Alice's adventure in wonderland[J/OL]. Digital scholarship in the humanities, 2022[2022 – 08 – 18]. https://doi. org/10. 1093/llc/fqac024.
[63] LIM T Y, LEE S H, TAK J Y, et al. Revisiting sylvia Plath's and Anne sexton's confessional poetry: analyzing stylistic differences and evolution of poetic voice (s) through computational text analysis[J]. Digital scholarship in the humanities, 2021, 36(4): 950 – 970.
[64] ALGEE-HEWITT M A. The hidden dictionary: text mining eighteenth-century knowledge networks[C/OL]. Mexico City: DH. 2018[2022 – 08 – 18]. https://dh2018. adho. org/en/the-hidden-dictionary-text-mining-eighteenth-century-knowledge-networks/.
[65] ZHITOMIRSKY-GEFFET M, PREBOR G. Sagebook: toward a cross-generational social network for the Jewish sages' prosopography[J]. Digital scholarship in the humanities, 2019, 34(3): 676 – 695.
[66] RUEGG C, LEE J J. Epic social networks and eve's centrality in Milton's paradise lost[J]. Digital scholarship in the humanities,

2020, 35(1): 146－159.
[67] KUBIS M, MICKIEWICZ A. Quantitative analysis of character networks in polish xix and xx century novels[C/OL]. Utrecht: DH, 2019[2022－08－18]. https://mkubis. home. amu. edu. pl/dh2019. pdf.
[68] LEE C. Do language combinations affect translators' stylistic visibility in translated texts? [J]. Digital scholarship in the humanities, 2018, 33(3): 592－603.
[69] LI W. Visualizing networks of artistic ideas in history paintings in the seventeenth-century Netherlands[C/OL]. Utrecht: DH, 2019[2022－08－18]. https://dataverse. nl/dataset. xhtml? persistentId=doi:10. 34894/YVFEZO.
[70] HAIDER T, VAN DYCK-HEMMING A. Extracting a social network of musicologists[C]// Digital Humanities 2020 Book of Abstracts. Ottawa: DH, 2020:143.
[71] LA PARRA-PEREZ A, MUÑOZ F F, FERNANDEZ-DE-PINEDO N. Econhist: a relational database for analyzing the evolution of economic history (1980－2019)[J]. Historical methods: a journal of quantitative and interdisciplinary history, 2022, 55(1): 45－60.
[72] BROWN D M, SOTO-COROMINAS A, SUÁREZ J L. The preliminaries project: geography, networks, and publication in the Spanish golden age[J]. Digital scholarship in the humanities, 2017, 32(4): 709－732.
[73] LEDOLTER J, VANDERVELDE L. A case study in text mining: textual analysis of the territorial papers[J]. Digital scholarship in the humanities, 2020, 35(1): 101－126.
[74] LEVINE M. Post WWI Chinese revolutionary leaders in Europe[J/OL]. Journal of historical network research, 2021, 5(1) [2022－08－18]. http://jhnr. uni. lu/index. php/jhnr/article/view/125.
[75] 阳雨,郭勇,李海龙,等. 异质网络社区发现研究进展[J]. 计算机应用研究,2018,35(10):2881－2887.

作者贡献说明：

王林旭：研究资料收集与分析，撰写论文；
王军：研究设计，修订论文。

Application Framework of Network Analysis in Digital Humanities

Wang Linxu　Wang Jun
Department of Information Management, Peking University, Beijing 100871
The Center for Digital Humanities, Peking University, Beijing 100871

Abstract: [**Purpose/Significance**] Network analysis is an effective analysis method and visualization method, is one of the most widely used directions in the field of digital humanities. A systematic summary and induction of the application of network analysis in the field of digital humanities will help the digital humanities researchers quickly identify the capabilities and limitations of network analysis, and to carry out deeper research practice. [**Method/Process**] In this study, we adopted content analysis to summary the papers published in the most influential journals in the international digital humanities field and International Digital Humanities Conference in the past five years. And from the five perspectives of research question, dataset, network special characteristics and network analysis metrics, the organization system for the application of network analysis methods in the field of digital humanities was finally extracted. [**Result/Conclusion**] The organizational structure includes three kinds of data scale (including single text analysis, parallel text analysis and corpus analysis), five application scenarios (containing characters analysis network, people relation network, words semantic network, text relevancy network and culture topic network), two scales of indicators (including global indicators and local indicators), and five types of metrics (like composition, density, centrality, clique and structure).

Keywords: network analysis　digital humanities　social network analysis　content analysis